Data Analytics Portfolio

Shirin Younesian

sh.younesian@gmail.com

github.com/shyounesian

Data Analysis Projects

Game Co
Video Game Popularity Data Project

Medical Staffing Agency
Preparing for Influenza Season

Storyboard in Tableau

Rockbuster Stealth
Launch an online video rental service

PostgreSQL Project in GitHub

Data Analysis Projects

Instacart Grocery

Instacart Grocery Basket Analysis

Python Project in GitHub

Global Super Store

Online Sales Trend Analysis

Time Series, Linear Regression & Clustering in Python in GitHub

Storyboard in Tableau

Data Analysis Tools

Excel

PowerPoint

Tableau

PostgreSQL

DBVisualiser

Python Libs
- Pandas
- Numpy
- Matplotlib
- OS
- Seaborn
- Sklearn
- Sklearn.cluster

Game Co

Project Name

Video Game Popularity Data Project

Object

Feasibility of developing and expanding sales of video games

Data

Historical sales of video games from 1980-2016 Source: VGCHartz

Skills

Applying Excel common features / Using pivot table / Creating visualizations in Excel / Cleaning Data / Descriptive Analysis / Storytelling with PowerPoint

Game Co Analysis
Regional sales form 1996- 2008

- ❑ In this period of time, we see an increase in sales in all regions, and in 2008, we experience the highest amount of sales
- ❑ NA with Ave. 50.69% (proportion of global sales) has the highest share
- ❑ EU with Ave. 26.3% is in second rank (Gradual growth of the market)
- ❑ JP with Ave. 15.53% is in third rank (Decrease in market share)
- ❑ Other with Ave. 7.5% is in fourth rank
- ❑ NA and JP experience more severe fluctuations compared to EU

Regional_Sales by Year
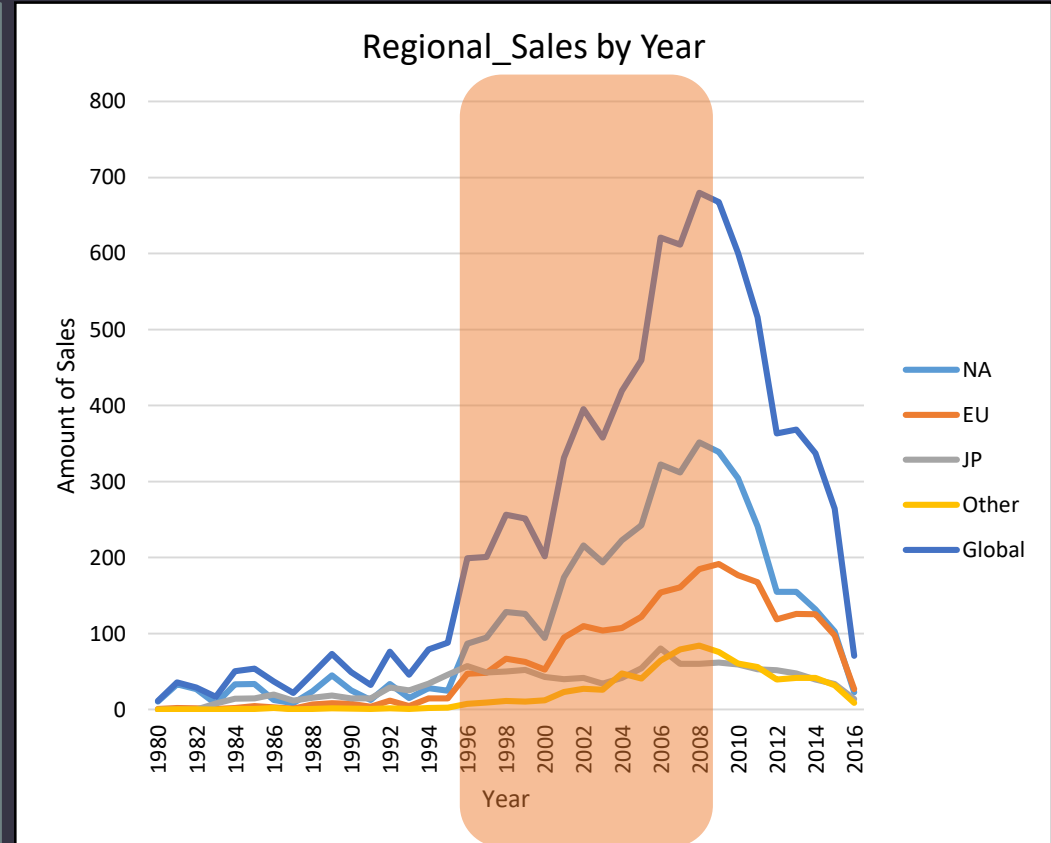
Game Co Analysis
Global and Regional sales by Genre

At the global level,
- ❑ The most sales are dedicated to Action, Sports & Shooter
- ❑ The lowest sales are dedicated to Strategy, Adventure & Puzzle
- ❑ JP has the most share in Role-Playing genre
- ❑ Popular genres in JP are different from all other regions

Global sales by genre

Regional_Sales by Genre
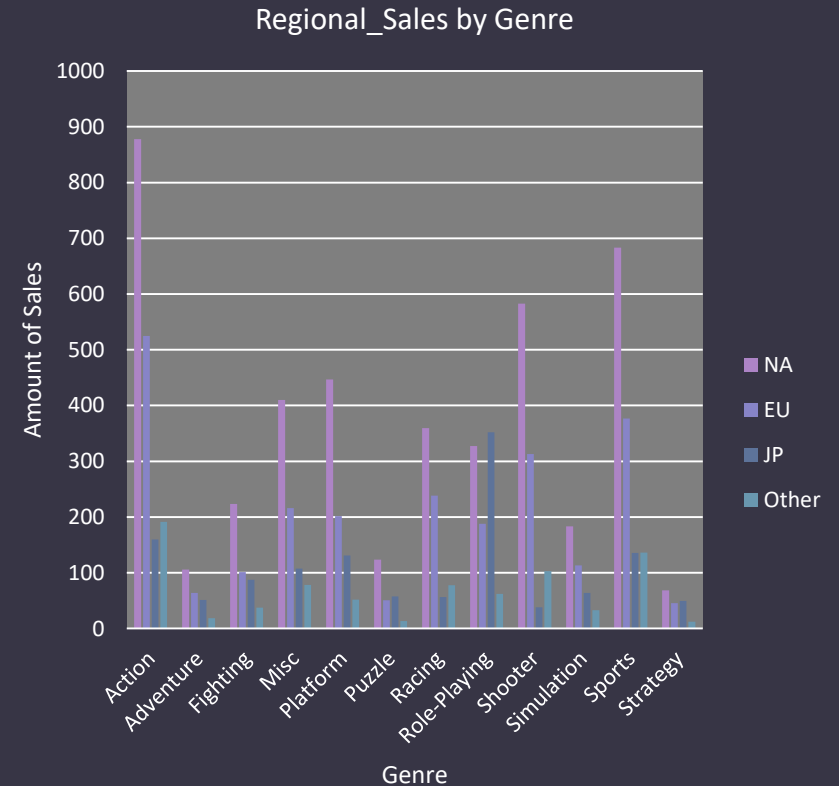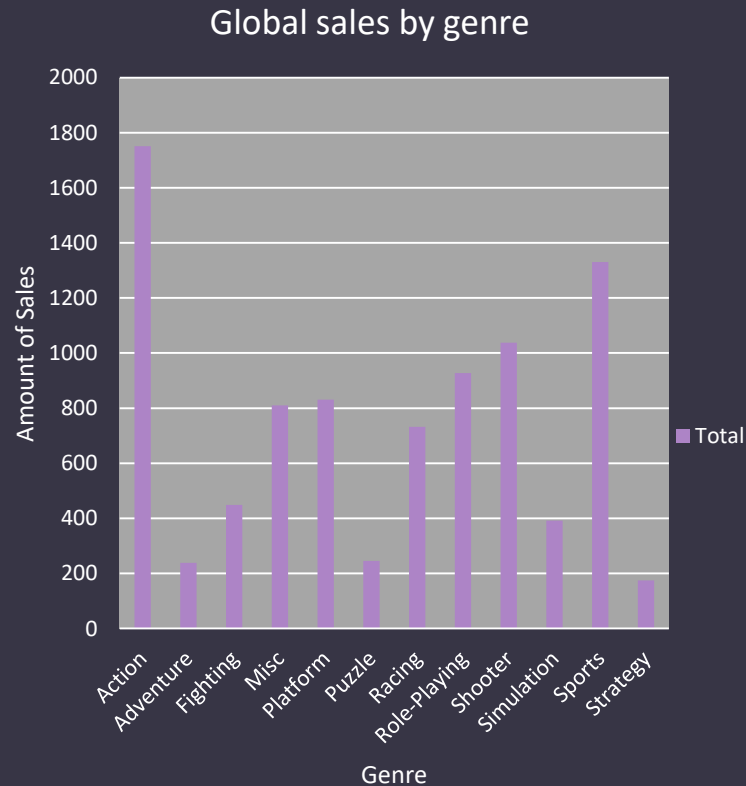
Game Co
Project Conclusion & Recommendation

- ❑ NA has played the largest and most influential role in global sales over the time period under our analysis

- ❑ NA has always been the leader in testing and experiencing the innovations of this market

- ❑ It can be concluded that the NA region should still be at the forefront of attention and it has a high sales potential

- ❑ Europe has experienced less fluctuations compared to NA and JP, and this is a sign of the relative stability of this market

- ❑ It can be concluded that the EU region can be considered as a strategic sales region for 2017

- ❑ JP has experienced a sales rate of 13% in 2015 and 19% in 2016, and this represents a growth of 6% in one year, which is a significant figure and indicates the possibility of market growth in 2017

- ❑ Considering the one-year growth of 6%, we can hope that marketing in this area can have a positive result

Medical Staffing Agency

Project Name

Preparing for Influenza Season

Object

During flu season, determine when to send medical staff and how many staff for each state.

Data

Influenza deaths, Source: CDC / Population data , Source: US Census Bureau

Skills

Statistical visualization , Spatial Analysis, Storytelling with Data Presentations in Tableau / Data Quality Checks(Integrity & consistency), Data Profiling, Statistical Analysis, Hypothesis Testing,

Storyboard in Tableau

Medical Staffing Agency Analysis
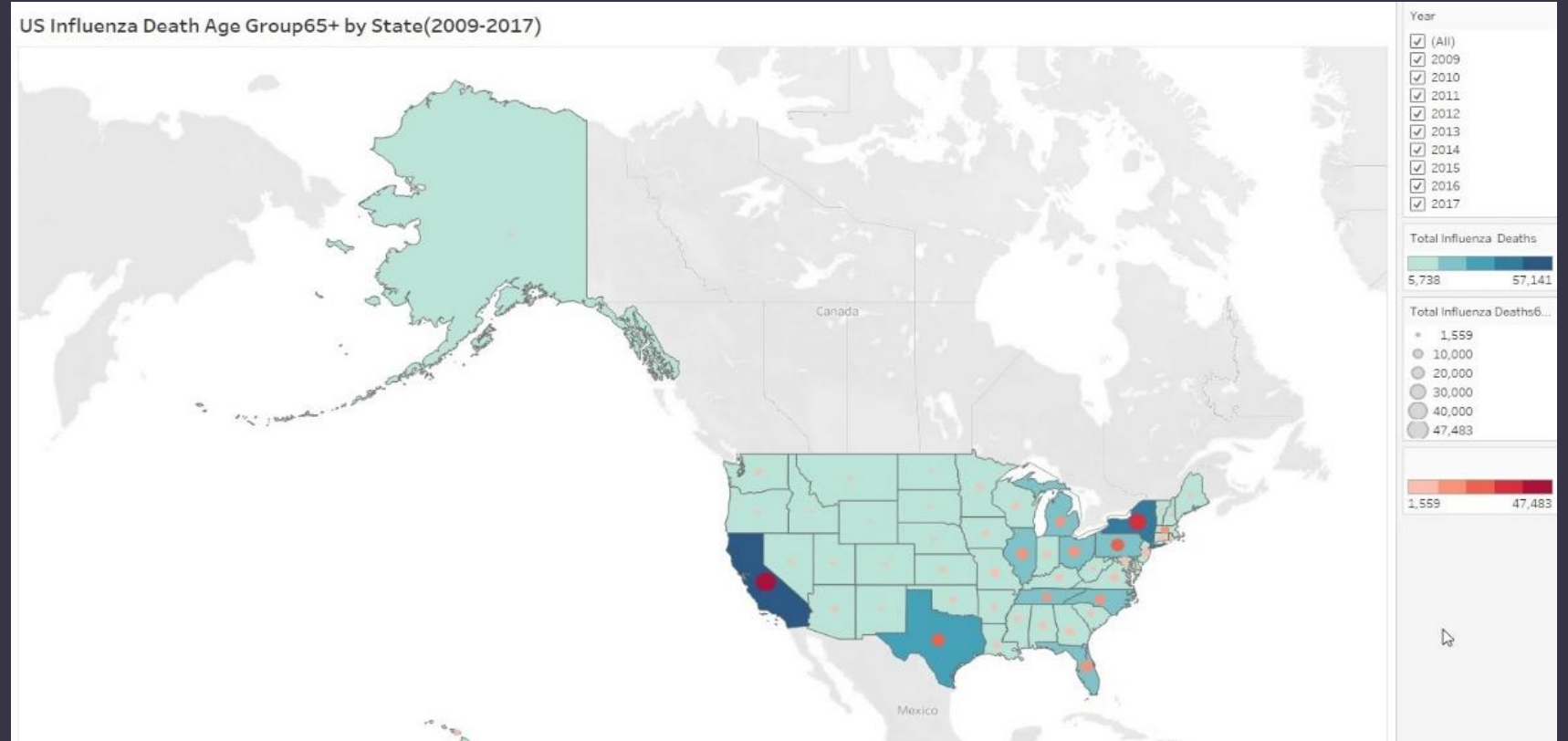## Influenza Deaths (actual & forecast) by Region

❑ The highest mortality is related to the South region and the lowest is related to the Northeast

❑ Most deaths occur in January, February, and March

❑ In some states, there is more than one flu season

US Influenza Deaths(actual & forecast) by Region 2009-2018
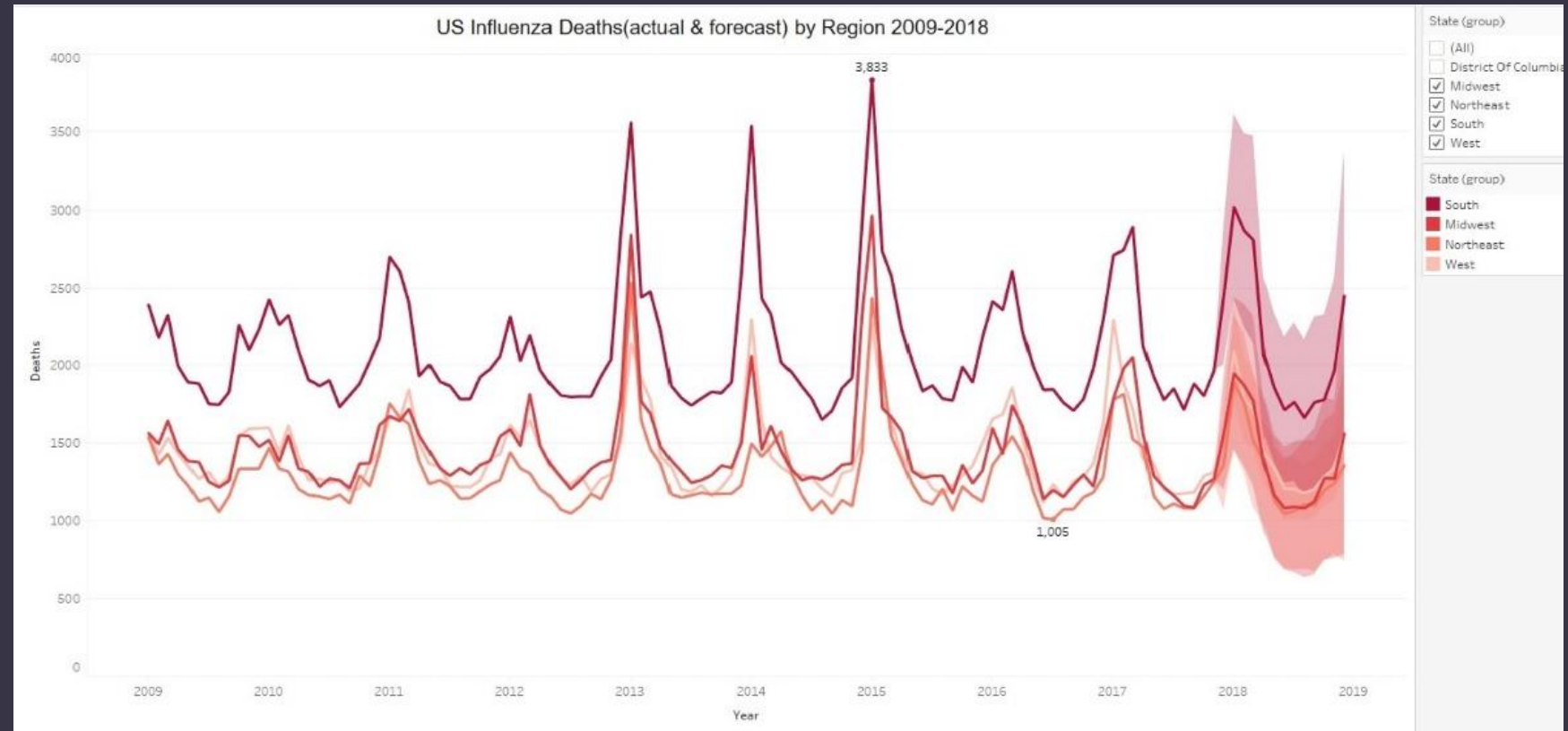
Medical Staffing Agency Analysis
Influenza Deaths by State

❑ California, New York & Texas, have the highest death rates.

❑ Alaska, Vermont & North Dakota, have the lowest death rates

US Influenza Death Age Group65+ by State(2009-2017)

Year
- ☑ (All)
- ☑ 2009
- ☑ 2010
- ☑ 2011
- ☑ 2012
- ☑ 2013
- ☑ 2014
- ☑ 2015
- ☑ 2016
- ☑ 2017

Total Influenza Deaths
5,738   57,141

Total Influenza Deaths6...
- · 1,559
- ○ 10,000
- ○ 20,000
- ○ 30,000
- ○ 40,000
- ○ 47,483

1,559   47,483

Medical Staffing Agency Analysis
Project Conclusion & Recommendation

❑ Human Factor

The vulnerable population of the age group of 65 years and over is more at risk of contracting the flu and dying. Therefore, states with a higher percentage of vulnerable populations should be prioritized for dispatching medical staff.

❑ Location Factor

The spread and death rate of influenza in southern and eastern states is more than in other regions. But during the years 2009 to 2017, California, New York, Texas, Pennsylvania and Florida have always had the highest death rates.

❑ Temporal Factor

Influenza outbreaks are observed in January, February, March, and December. Some states have more than one flu season.

Rockbuster Stealth

Project Name

Launch an online video rental service

Object

Facing stiff competition from streaming services such as Netflix, the management is planning to launch an online video rental service in order to stay competitive.
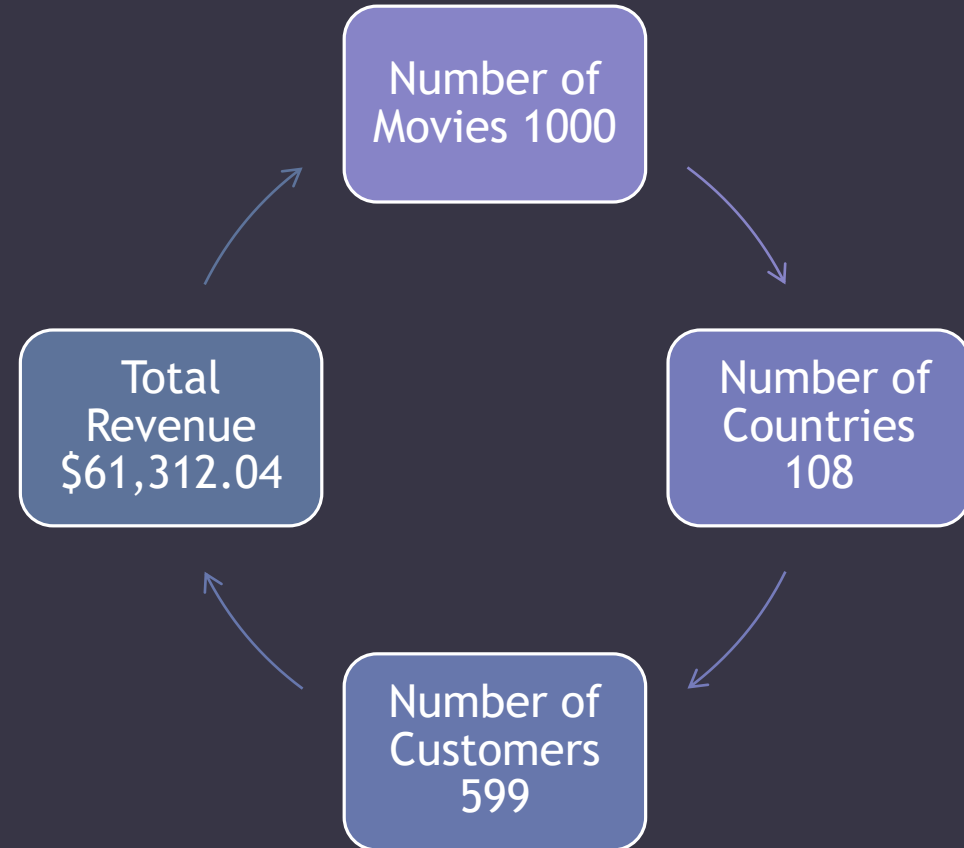
Data

Rockbuster Data Set

Skills

Extract Entity Relationship Diagram(ERD), Data Dictionary, OLAP Schemas, SQL Commands & Queries (CRUD, Join, Subquery, …), Cleaning & Summarizing data in SQL

PostgreSQL Project in GitHub

Rockbuster Stealth Analysis
Statistical Information

Rental Duration (day)	Rental Rate ($/day)	Replacment Cost ($)
• MAX 7 • AVG 5 • MIN 3	• MAX 5 • AVG 3 • MIN 1	• MAX 30 • AVG 20 • MIN 10

Number of Movies 1000

Number of Countries 108

Number of Customers 599

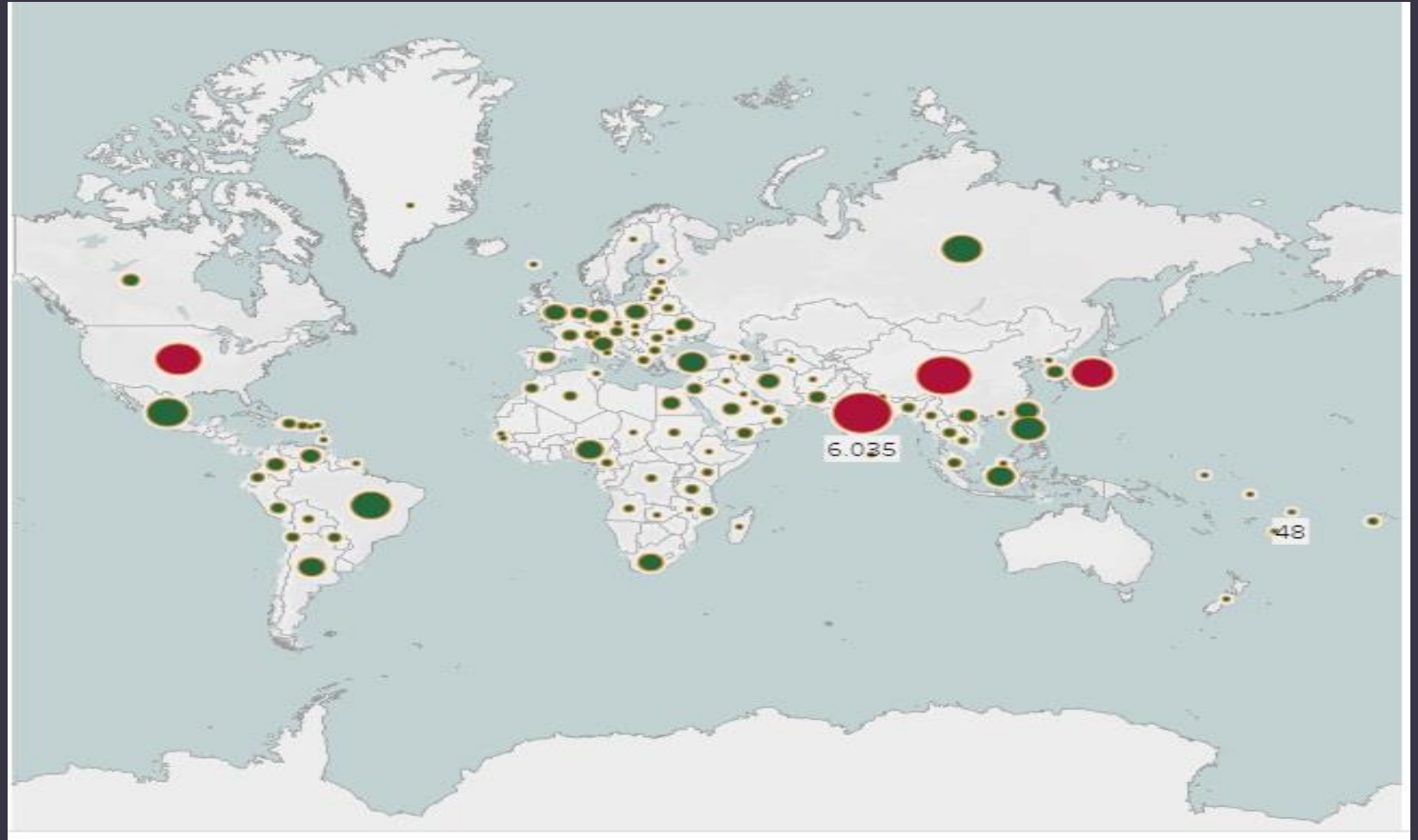Total Revenue $61,312.04

Rockbuster Stealth Analysis
Geographical Distribution of Customers and the Revenue

❑ Based on the two factors of customer and revenue, the top countries among 108 countries are:

- ▪ India
- ▪ China
- ▪ Japan
- ▪ United States

❑ American Samoa with 1 customer and $48 payment is the lowest income region

Link to Public Tableau

Rockbuster Stealth
Project Conclusion & Recommendation

❑ Data analysis shows that movies with PG_13 rating, have the highest rental share and have the most demand for renting, and movies with G rating, have the lowest rental share.

❑ The most popular genres are Sport, Animation, and Action, which have the highest movie rental rates and the lowest demand is for the genre Thriller.

❑ From a regional point of view, Asia has the highest amount of revenue and the number of customers, followed by America

❑ Attention to the cultural and taste differences and the language of customers (in all age groups)in each country

❑ Attention to the limitations in technology and communication infrastructure in each country

❑ Feasibility of local payment solutions and, Income level of customers in different countries

Instacart Grocery

Project Name

Instacart Grocery Basket Analysis

Object

Identify sales patterns and derive insights to suggest strategies for better sales and services.

Data

Instacart Data Sets

Skills

Data Wrangling & Subsetting, Data Consistency Checks, Combining & Exporting Data, Visualization with Python

Python Project in GitHub

Instacart Grocery Analysis
Order and Sales amount by the days of week

- ❑ Saturday and Sunday have the highest number of orders

- ❑ Saturday and Friday have the best sales

Spending habits by days of week
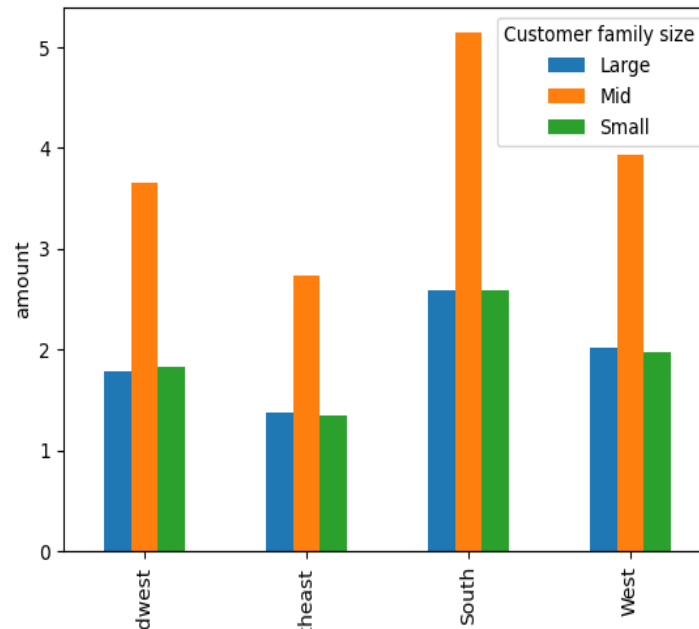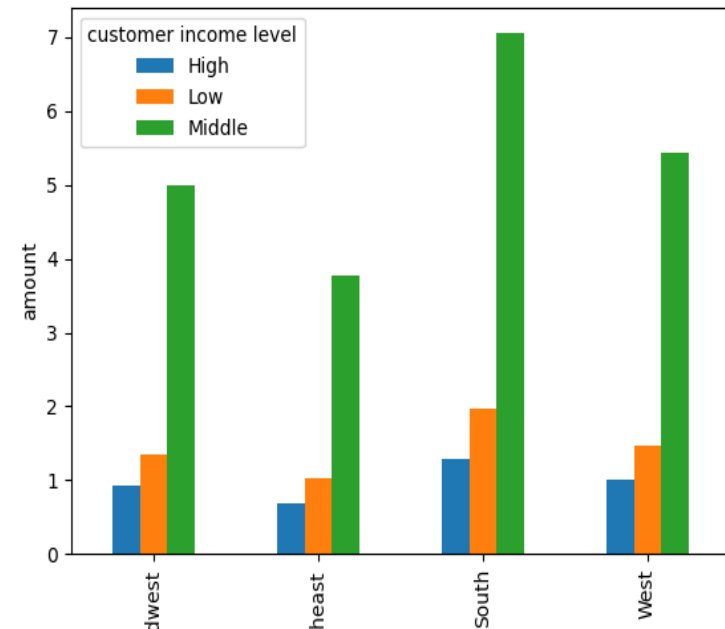

Order day of week
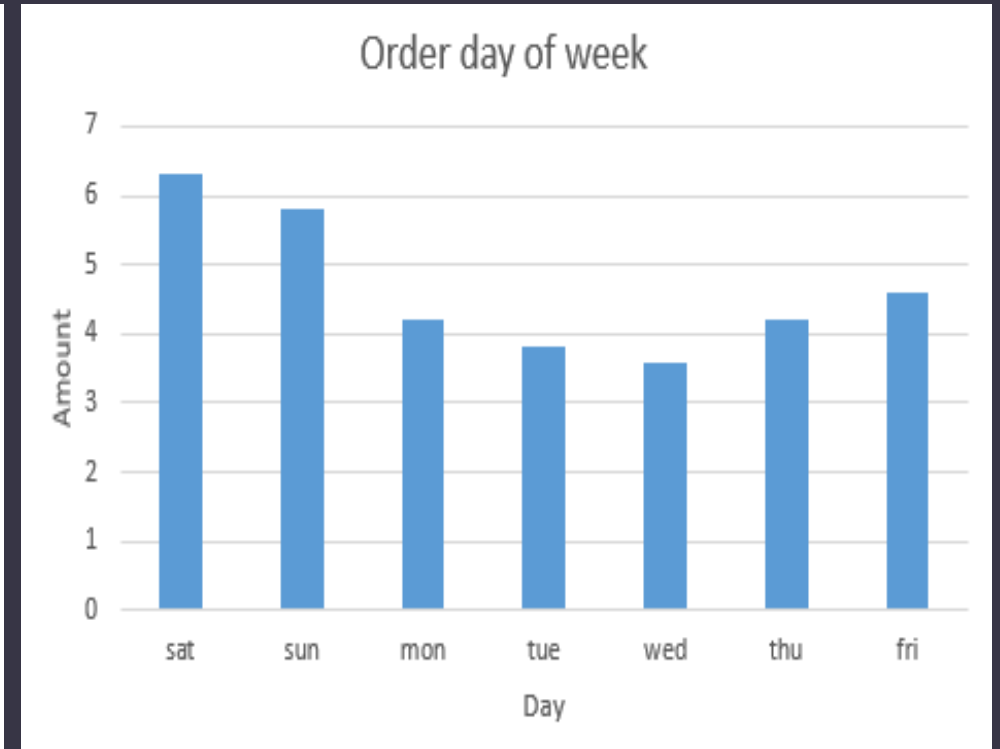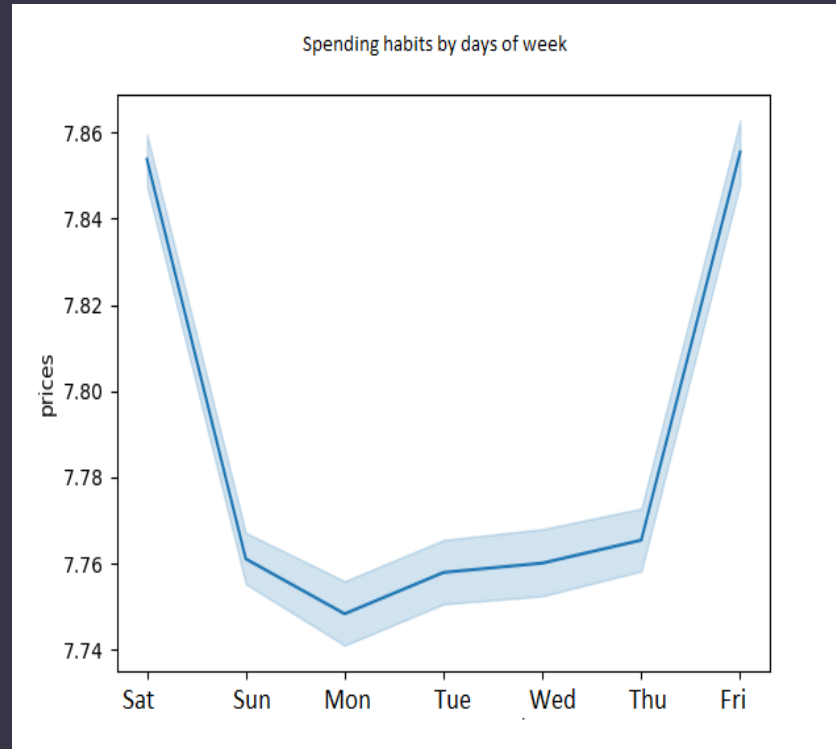
Instacart Grocery Analysis
Customer distribution in regions by Income & Family Size

- Middle-income customers (60k up to 150k) have the highest rank in all regions

- Medium-sized family customers(2 up to 4 people) have the highest rank in all regions

Instacart Grocery
Project Conclusion & Recommendation

❑ From the point of view of the geographical regions, the southern region and then the western region have the highest number of order registrations. Therefore, more warehouses can be considered for these areas so that orders can be received faster.

❑ Customers with average income and dependents have the highest number of purchases and orders. Usually, these customers are middle-aged, followed by single customers. It is suggested that the taste of these groups should be given more attention. On the other hand, incentive policies can be considered for them.

❑ It is suggested that different tastes and demands should be considered, and the company should maintain the diversity of its products or increase them.

❑ The price, it was found that the products with average price are the most welcomed and ordered by the customers. However, customers with low-income levels should be kept by offering more affordable products, and wealthy customers by offering special and luxury products.

Global Super Store

Project Name

Online Sales Performance Analysis

Object

Identifying sales trends and discovering the relationships of sales related variables to predict sales behavior.

Data

Global Super Store dataset

Skills

Correlation Matrix, Linear Regression, Time Series, k-means and Clustering Analysis in Python

Time Series, Linear Regression & Clustering in Python in GitHub

Storyboard in Tableau
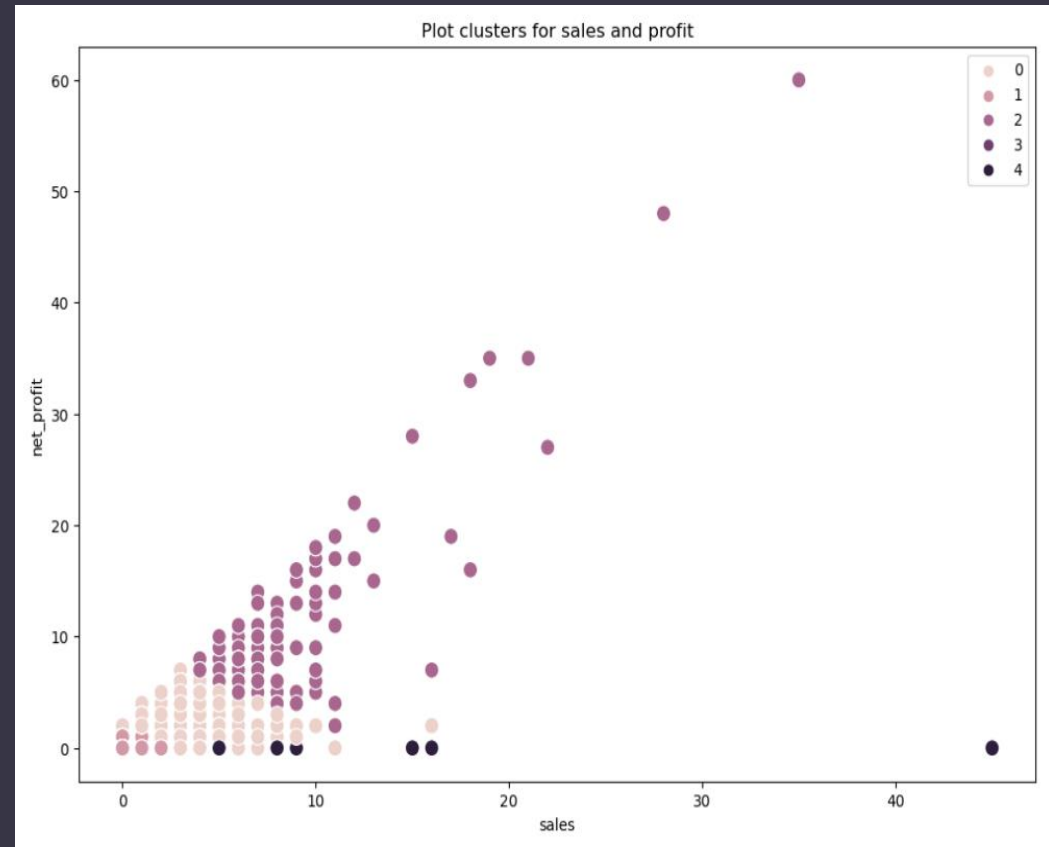
Global Super Store
Variables Relationships Analysis

☐ In this Correlation Matrix, we can see the relationship between the variables at a glance.

☐ The number inserted in each cell shows the correlation coefficient between the two variables.

☐ Negative coefficients indicate that with the increase of one variable, the second variable decreases.

☐ Positive coefficients indicate that with the increase of one variable, the second variable also increases.

Correlation Matrix
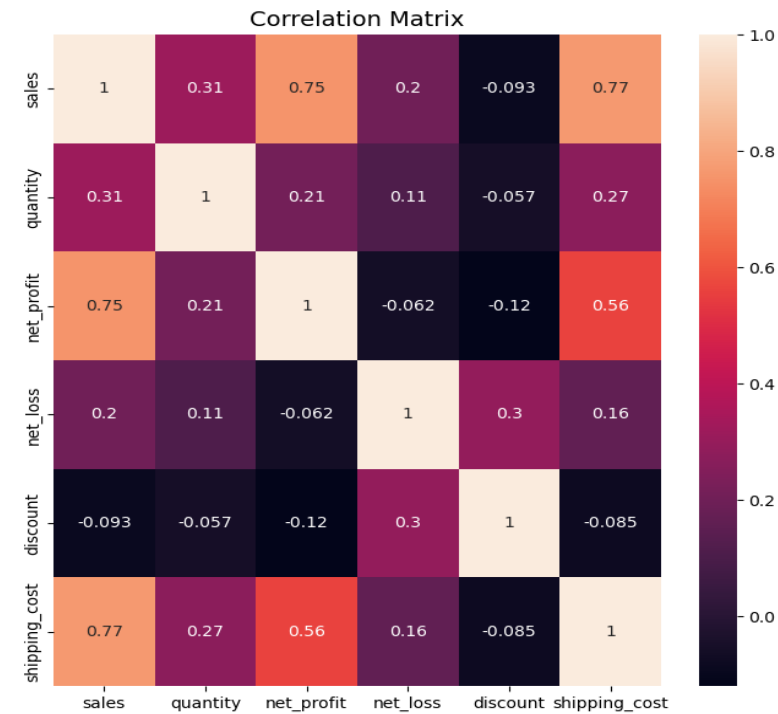
Global Super Store
Unsupervised Clustering Analysis

❑ Based on the analysis, we found that the most suitable number of clusters (k-means) is 5.

❑ In the clustering of sales and profit, the maximum profit (here 60 units) belongs to sales points with a maximum of 35 units and a median of 8 and a mean of 9 (points with purple color).

❑ In fact, as can be seen in the graph, sales between 4 and 35 units have the highest profit margin. The distribution of purple points is also in the same range.

❑ In sales of less than 3 units or more than 35 units, the profit margin is zero.


Plot clusters for sales and profit

Global Super Store
Time Series Analysis

- The first line is the actual data, which is called the level.

- The second is the underlying trend, the overall trend continues to increase all the way until the half of 2014.

- The third component is seasonality. Here, we can see seasonal fluctuations that change at regular intervals (as we can see 6th and 12th months of each year have the highest sales and then gradually decrease).

- The fourth is the noise (residual)

- Based on the above results the time series was non-stationary, so we couldn't forecast with non-stationary data, so at first, we Stationarized the dataset and then we were able to plot the actual and estimated situation.

Decomposition Sales Time Series
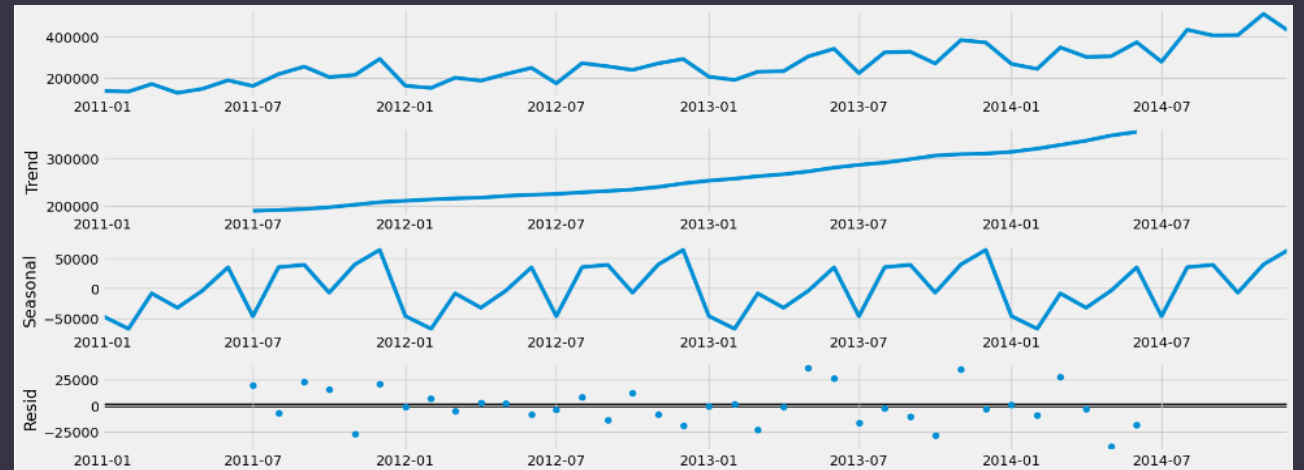
Actual and Estimate Sales Analysis

Global Super Store
Project Conclusion & Recommendation

❑ Temporal Factor

Based on the results of the data analysis, it shows that an internal pattern can be seen based on sales fluctuations, which states that the months of June and December have the highest amount of sales in each year and the lowest sales in the month of January and July have been recorded. On the other hand, based on the sales forecast for 2015, this pattern is still observed. So the store should prepare itself for periods of recession and prosperity.

❑ Geographical Factor

The United States (with 18% of total sales), Australia (7%), and France (5%) have the highest sales, the highest number of customers, and the highest profits, respectively, among the 143 countries in the dataset. Therefore, special strategies can be considered for top countries. Of course, to expand its field of activity, it can also count on China, which is among the top 10 countries in global sales and has a high potential for sales growth and attracting more customers.

❑ Sales Range

Based on the results of sales clustering, we found that sales in all ranges do not have high profits and sometimes even cause losses. Therefore, based on the findings of the analysis, sales of more than 500 and less than 8000 have the highest amount of profit. Fewer sales usually result in lower profit margins and usually face losses.

Next Project

COMING SOON

THANK YOU

Shirin Younesian

sh.younesian@gmail.com

github.com/shyounesian